

# Summer 2022 Data Science Intern Challenge

GitHub : [Shopify Technical Challenge Submission \(link\)](#)

**Question 1:** Given some sample data, write a program to answer the following: [click here to access the required data set](#)

On Shopify, we have exactly 100 sneaker shops, and each of these shops sells only one model of shoe. We want to do some analysis of the average order value (AOV). When we look at orders data over a 30 day window, we naively calculate an AOV of \$3145.13. Given that we know these shops are selling sneakers, a relatively affordable item, something seems wrong with our analysis.

- a. Think about what could be going wrong with our calculation. Think about a better way to evaluate this data.

Answer: I have used Python, Pandas and Jupyter noitebook to automate the AOV calculation.

Step 1 : Read the csv file using `read.csv()`

Step 2 : convert it into dataframe using `.DataFrame()` for better visualization

Step 3 : Find summation for 2 metrics individually using `.sum()`

Step 4 : calculate AOV

- b. What metric would you report for this dataset?

Answer : The metrics we need to report this dataset are : `'order_amount'.sum()` and `'total_amount'.sum()`

- c. What is its value?

Answer : The value is 357.921 dollar

**Question 2:** For this question you'll need to use SQL. [Follow this link](#) to access the data set required for the challenge. Please use queries to answer the following questions. Paste your queries along with your final numerical answers below.

- a. How many orders were shipped by Speedy Express in total?

Answer : 54

```
SELECT COUNT(*) as Total_SpeedyExpress_Order
FROM Shippers
LEFT JOIN Orders
ON Shippers.ShipperID = Orders.ShipperID
WHERE ShipperName = "Speedy Express"
```

- b. What is the last name of the employee with the most orders?

**Answer: Peacock**

```
SELECT Employees.LastName, COUNT(DISTINCT Orders.OrderID) as  
most_no_of_orders  
FROM Employees  
LEFT JOIN Orders  
ON Employees.EmployeeID = Orders.EmployeeID  
GROUP BY Employees.EmployeeID  
ORDER BY most_no_of_orders DESC  
LIMIT 1
```

- c. What product was ordered the most by customers in Germany?

**Answer: "Boston Crab Meat" (Total Item of 60)**

```
SELECT Products.ProductName, OrderDetails.OrderID, SUM(OrderDetails.Quantity) as  
Total_ITEM, Orders.CustomerID, Customers.Country  
FROM Products  
LEFT JOIN OrderDetails  
ON Products.ProductID = OrderDetails.ProductID  
LEFT JOIN Orders  
ON OrderDetails.OrderID =Orders.OrderID  
LEFT JOIN Customers  
ON Orders.CustomerID = Customers.CustomerID  
WHERE Customers.Country = "Germany"  
GROUP BY Products.ProductName  
ORDER BY Total_ITEM DESC  
LIMIT 1
```